Classification of cancer

Field of invention

The present invention relates to a method for classification of cancer in an individual, wherein the microsatellite status and a prognostic marker are determined by examining gene expression patterns. The invention also relates to various methods of treatment of cancer. Additionally, the present invention concerns a pharmaceutical composition for treatment of cancer and uses of the present invention. The invention also relates to an assay for classification of cancer.

Background of invention

Studies of differential gene expression in diseased and normal tissues have been greatly facilitated by the building of large databases of the human genome sequences. Gene expression alterations are important factors in the progression from normal tissue to diseased tissue. In order to obtain a profile of transcriptional status in a certain cell type or tissue, array-based screening of thousands of genes simultaneously is an invaluable tool. Array-based screening even allows for the identification of key genes that alone, or in combination with other genes, regulate the behaviour of a cell or tissue. Candidate genes for future therapeutic intervention may thus also be identified.

Colorectal cancer generally occurs in 1 out of every 20 individuals at some point during their lifetime. In the United States alone about 150,000 new cases are diagnosed each year which amount to 15% of the total number of new cancer diagnoses. Unfortunately, colorectal cancer causes about 56,000 deaths a year in the United States.

The malignant transformation from normal tissue to cancer is believed to be a multistep process. Two molecular pathways are known to be involved in the development of colorectal cancer (Lengauer C, Kinzler KW, Vogelstein B., 1998) namely the microsatellite stable (MSS) pathway and the microsatellite instable (MSI) pathway. MSS is associated with high frequency of allelic losses, abnormalities of cytogenetic nature and abnormal tumor content of DNA. MSI however is associated with defects in the DNA mismatch repair system which leads to increased rate of point mutations and minor chromosomal insertions or deletions.

MSI tumors can be of hereditary or sporadic nature. Ninety percent of MSI tumours are of sporadic origin. Sporadic tumours are presumably MSI due to epigenetic hypermethylation of the MLH1 gene promoter. The hereditary tumours account for 10 % of the MSI tumors. Mutations of for example the MLH1 or MSH 2 genes are often the cause of hereditary tumor development.

The ability of being able to determine the sporadic or hereditary nature of a MSI tumor is highly valuable. In case a tumor is characterized as being MSI , and certain clinical criteria are fulfilled such as age below 50 or three first degree relatives with colon cancer, a screening programme of family members for early diagnosis and treatment of potential colon or endometrial cancer development is initiated. The human and economic costs in relation to screening programmes are severe. Consequently, a need for identifying colon cancers with a hereditary character exists. Further, these patients have a poor prognosis, as they have an increased risk of metachronous colon tumors and a highly increased risk of getting cancer in the endometrium (females), upper urinary tract and a number of other organs. Thus, one may regard the determination of a colon tumor as being sporadic or hereditary as determination of a prognostic factor.

Tumors appearing to be similar – morphologically, histochemically or microscopically – can be profoundly different. They can have different invasive and metastasizing properties, as well as respond differently to therapy. There is thus a need in the art for methods which distinguish tumors and tissues on different bases than are currently in use in the clinic. Determination of microsatellite status using an array-based methodology is faster than conventional DNA based methods, as it does not require microdissection, and forms a set of genes that can be combined with other sets of genes on a colon cancer array that can be used to determine microsatellite status as well as e.g. predict disease course by identifying hereditary cases or other prognostic important factors, and finally predict therapy response.

Summary of invention

In one aspect the present invention relates to a method of classifying cancer in an individual having contracted cancer comprising

in a sample from the individual having contracted cancer determining the microsatellite status of the tumor and

in a sample from the individual having contracted cancer, said sample comprising a plurality of gene expression products the presence and/or amount which forms a pattern, determining from said pattern a prognostic marker, wherein the microsatellite status and the prognostic marker is determined simultaneously or sequentially

classifying said cancer from the microsatellite status and the prognostic marker.

The cancer may be any cancer known to be microsatellite instable in at least a fraction of the cases, such as colon cancer, uterine cancer, ovary cancer, stomach cancer, cancer in the small intestine, cancer in the biliary system, urinary tract cancer, brain cancer or skin cancer. These cancers are part of the spectrum of cancers that belong to the hereditary non-polyposis colon cancer syndrome, but the invention is not limited to this syndrome.

Gene expression patterns may be formed by only a few genes, but it is also a preferred embodiment that a multiplicity of genes form the expression pattern whereby information for classification of cancer can be obtained.

Furthermore, the invention relates to a method for classification of cancer in an individual having contracted cancer, wherein the microsatellite status is determined by a method comprising the steps of

in a sample from the individual having contracted cancer, said sample comprising a plurality of gene expression products the presence and/or amount of which forms a pattern that is indicative of the microsatellite status of said cancer,

determining the presence and/or amount of said gene expression products forming said pattern,

obtaining an indication of the microsatellite status of said cancer in the individual based on the step above.

Yet another aspect of the invention relates to a method for classification cancer in an individual having contracted cancer, wherein the hereditary or sporadic nature is determined by a method comprising the steps of

in a sample from the individual having contracted cancer, said sample comprising a plurality of gene expression products the presence and/or amount of which forms a pattern that is indicative of the hereditary or sporadic nature of said cancer,

determining the presence and/or amount of said gene expression products forming said pattern,

obtaining an indication of the hereditary or sporadic nature of said cancer in the individual based on the step above.

The present invention further concerns a method for treatment of an individual comprising the steps of

selecting an individual having contracted a colon cancer, wherein the microsatellite status is stable, determined according to any of the methods as defined herein

treating the individual with anti cancer drugs .

Another aspect of the present invention relates to a method for treatment of an individual comprising the steps of

selecting an individual having contracted a colon cancer, wherein the microsatellite status is instable, determined according to any of the methods as defined herein

treating the individual with anti cancer drugs.

Yet another aspect of the present invention relates to a method for reducing malignancy of a cell, said method comprising

contacting a tumor cell in question with at least one peptide expressed by at least one gene selected from genes being expressed at least two-fold higher in tumor cells than the amount expressed in said tumor cell in question.

Additionally, the present invention concerns a method for reducing malignancy of a tumor cell in question comprising,

obtaining at least one gene selected from genes being expressed at least two fold lower in tumor cells than the amount expressed in normal cells

introducing said at least one gene into the tumor cell in question in a manner allowing expression of said gene(s).

The invention also relates to a method for reducing malignancy of a cell in question, said method comprising

obtaining at least one nucleotide probe capable of hybridising with at least one gene of a tumor cell in question, said at least one gene being selected from genes being expressed in an amount at least two-fold higher in tumor cells than the amount expressed in normal cells, and

introducing said at least one nucleotide probe into the tumor cell in question in a manner allowing the probe to hybridise to the at least one gene, thereby inhibiting expression of said at least one gene.

In a further aspect the invention relates to a method for producing antibodies against an expression product of a cell from a biological tissue, said method comprising the steps of

obtaining expression product(s) from at least one gene said gene being expressed as defined herein

immunising a mammal with said expression product(s) obtaining antibodies against the expression product.

The present invention also concerns a method for treatment of an individual comprising the steps of

selecting an individual having contracted a colon cancer, wherein the microsatellite status is stable, determined according to any of the methods as defined herein

introducing at least one gene into the tumor cell in a manner allowing expression of said gene(s).

The present invention further relates to a pharmaceutical composition for the treatment of a classified cancer comprising at least one antibody as defined herein.

In yet another aspect the invention concerns a pharmaceutical composition for the treatment of a classified cancer comprising at least one polypeptide as defined herein.

Further, the invention relates to a pharmaceutical composition for the treatment of a classified cancer comprising at least one nucleic acid and/or probe as defined herein.

In an additional aspect the present invention relates to an assay for classification of cancer in an individual having contracted cancer, comprising

at least one marker capable of determining the microsatellite status in a sample and

at least one marker in a sample determining the prognostic marker, wherein the microsatellite status and the prognostic marker is determined simultaneously or sequentially.

Detailed description of the drawings

Figure 1

Unsupervised hierarchical clustering of colorectal tumors based on the 1239 genes with the highest variation across all tumors.

The phylogenetic tree shows the spontaneous clustering of tumor samples and normal biopsies. Germline mutation indicates samples with hereditary mutations in

either *MLH1* or *MSH2* genes. In columns referring to results of immunohistochemistry a plus indicates a positive antibody staining. Tumor location indicates right-sided or left-sided location in the colon of the tumor.

Figure 2

Summary of the performance of the microsatellite instability classifier based on microarray data.

Panel A shows the number of classification errors as a function of the number of genes used. Panel B shows $\log_2$ of the ratio of the distance between a tumor to the centers of the microsatellite instable group and the microsatellite stable tumors. A value of +2 indicates that the distance of a tumor to the microsatellite instable group is 4 times the distance to the microsatellite stable group. Open bars are MSI tumors and solid bars are MSS tumors. Panel C shows the result of the permutation analysis for estimation of the stability of the classifier. This was estimated by generating one hundred new classifiers based on randomly chosen datasets from the 101 tumors each consisting of 30 microsatellite stable and 25 microsatellite instable samples. In each case the classifier was tested with the remaining 46 samples. The performance for each set was evaluated and averaged over all 100 training and test sets.

Figure 3

Classification of MSI tumors as hereditary or sporadic cases based on two genes.

Panel A shows the number of classification errors as a function of the number of genes used. In crossvalidation we found a minimum number of one error using two genes and adding more genes increased the number of errors to a maximum number of twelve. Both genes were used in at least 36 of the 37 crossvalidation loops. Panel B shows $\log_2$ of the ratio of the distance between a tumor to the centers of the sporadic microsatellite instable group and the hereditary microsatellite instable group. Panel C shows microarray signal values for *MLH1* and *PIWIL1* genes for all tumors. Asterisk indicates the misclassified tumor

Figure 4

Classification of microsatellite-instability status based on real-time PCR.

Panel A shows a cluster analysis of 18 of the 101 tumors samples and 9 genes based on the microarray data and compared to real-time PCR data from same samples and genes. Dark colors indicate relative low expression and light/light grey color palette high expression. Panel B shows the result of 47 new independent samples based on PCR data from 7 of the 9 genes. Relative distances are explained in the legend to figure 2. The two misclassified tumors are indicated with an asterisk. For PCR primers and hybridization probes see supplement to methods.

Figure 5

Kaplan-Meier estimates of crude survival among patient with Stage II and Stage III colorectal cancer according to microsatellite status of the tumor, determined by gene expression. Open triangles indicate censored samples. The patients left at risk are denoted in brackets. The P values were calculated with use of the log-rank test.

Figure 6

Phylogenetic tree resulting from unsupervised hierarchical clustering. Clusteranalysis of colon specimens with associated clinicopathological features.

Figure 7

Multidimentional scaling plot showing distances between groups of tumors.

Figure 8

Performance of prediction of survival before and after separation in MSI-H and MSS

Figure 9

Performance of the classifier for identification of hereditary disease.

Figure 10

Kaplan Meier estimates of overall survival among patients with Dukes' B and Dukes' C colon cancer according to microsatellite-instability status of the tumor, determined by gene expression.

Detailed description of the invention

Classification of cancer

The present inventors have, using large-scale array-based screenings, found a pool of genes, the expression products of which may be used to classify cancer in an individual. The presence of expression products and level of expression products provides an expression pattern which is correlated to a specific status and/or prognostic marker of the cancer. Characterization of the genes or functional analysis of the gene expression products as such is not required to classify the cancer based on the present method. Thus, the expression products of the plurality of genes can be used as markers for the classification of disease.

One aspect of the present invention concerns a method for classifying cancer in an individual having contracted cancer by determining the microsatellite status and a prognostic marker in a sample. Determination of the microsatellite status and the prognostic marker may be performed simultaneously or sequentially. In one embodiment of the present invention the microsatellite status is determined. The prognostic marker is determined in a sample, wherein the presence and/or the amount of a number of gene expression products form a pattern wherefrom the prognostic marker is determined. Based on the information gathered from the microsatellite status and the prognostic marker the cancer can be classified. In a preferred embodiment the prognostic marker is the hereditary or sporadic nature of the cancer. The hereditary or sporadic nature of the cancer can be determined through a number of steps comprising determining the presence and/or amount of gene expression products forming a pattern in a sample. The sample comprises a number of gene expression products the presence and/or amount of which forms a pattern that is indicative of the hereditary or sporadic nature of the cancer. Hereby, an indication of the hereditary or sporadic nature of the cancer is obtained.

In one embodiment of the invention the microsatellite status is determined using conventional analysis of microsatellite status as described elsewhere herein.

In another embodiment of the present invention the microsatellite status is determined by gene expression patterns wherein the presence and/or the amount of the gene expression products form a pattern that is indicative of the microsatellite status.

Classification of cancer provides knowledge of the survival chances of an individual having contracted cancer. In case of cancer which according to the present invention has been classified as a hereditary cancer, screening programmes of family members to the individual having the classified cancer can be initiated. Such screening programmes can comprise conventional screening programmes employing sequencing and other methods as described elsewhere. Thus, individuals at risk of developing cancer may be identified and action taken accordingly to detect developing cancer at an early stage of the disease greatly improving the chances of successful intervention and thus survival rates.

Classification of cancer also provides insights on which sort of treatment should be offered to the individual having contracted cancer, thus providing an improved treatment response of the individual. Likewise, the individual may be spared treatment that is inefficient in treating the particular class of cancer and thus spare the individual severe side effects associated with treatment that may even not be suitable for the class of cancer.

Microsatellite status

The use of highly variable repetitive sequences found in microsatellite regions adjacent to genes or other areas of interest may be used as markers for linkage analysis, DNA fingerprinting, or other diagnostic application.

Microsatellites are defined as loci (or regions within DNA sequences) where short sequences of DNA are repeated in tandem repeats. This means that the sequences are repeated one right after the other. The lengths of sequences used most often are di-, tri-, or tetra-nucleotides. At the same location within the genomic DNA the number of times the sequence (ex. AC) is repeated often varies between individuals, within populations, and/or between species. Due to the many repeats the microsatellites are prone to alter if there is a reduced repair of mismatches in the genome.In the present invention the traditional method of determining microsatellite status by employing microsatellite markers is replaced by determination of gene expression patterns.

An important factor in multi-step carcinogenesis is genomic instability. The development of some cancer forms is known to follow two distinct molecular routes. One

route is the microsatellite stable, MSS, (and chromosomal instable pathway) which is often associated with a high frequency of allelic losses, cytogenetic abnormalities and abnormal DNA tumor contents. The second route is the microsatellite instable pathway MSI that is characterized by defects in the DNA mismatch repair system which leads to a high rate of point mutations and small chromosomal insertions and deletions. The small chromosomal insertions and deletions can be detected as mono and dinucleotide repeats (Boland CR, Thibodeau SN, Hamilton SR, et al., Cancer Res 1998;58(22):5248-57).

One aspect of the present invention relates to the classification of cancer in an individual having contracted cancer by determining the microsatellite status and a prognostic marker. One embodiment of the invention relates to microsatellite status determined by conventional methods employing microsatellite analysis as described above. Another embodiment of the invention relates to establishing the microsatellite status by determining the presence and/or amount of gene expression products of a sample which comprises a plurality of gene expression products forming a pattern which is indicative of the microsatellite status.

The expression products of genes according to the present invention are not necessarily identical to the genes that are analysed by microsatellite markers in conventional methods of determining microsatellite status. The pattern of the gene expression products according to the present invention however correlates with information on microsatellite status that can be obtained using traditional methods.

The determination of the microsatellite status and the prognostic marker of the cancer may be performed sequentially. However, the determinations may also be performed simultaneously.

Prognostic marker

Together with knowledge of the microsatellite status in a sample of an individual having contracted cancer a prognostic marker is employed for classifying the cancer. The prognostic marker may be any marker that provides knowledge of the cancer type when combined with knowledge of microsatellite status. Consequently the prognostic marker may provide additional information on the cancer type when the microsatellite status is stable and similarly when the microsatellite status is

instable. In a preferred embodiment of the present invention the prognostic marker is the hereditary or sporadic nature of a cancer given that the microsatellite status is instable. The prognostic marker may in another embodiment be a prognostic marker for any feature or trait that provides further possibilities of classifying cancer.

The prognostic marker is determined in a sample comprising a number of gene expression products wherein the presence and/or amounts of gene expression products form a pattern that is indicative of the prognostic marker.

Hereditary and sporadic nature of cancer

Hereditary nonpolyposis colon cancer (HNPCC) is a hereditary cancer syndrome which carries a very high risk of colon cancer and an above-normal risk of other cancers (uterus, ovary, stomach, small intestine, biliary system, urinary tract, brain, and skin). The HNPCC syndrome is due to mutation in a gene in the DNA mismatch repair system, usually the MLH1 or MSH2 gene or less often the MSH6 or PMS2 genes. Families with HNPCC account for about 5% of all cases of colon cancer and typically have the following features (called the Amsterdam clinical criteria):

Three or more first relative family members with colorectal cancer; affected family members in two or more generations; and at least one person with colon cancer diagnosed before the age of 50.

The highest risk with HNPCC is for colon cancer. A person with HNPCC has about an 80% lifetime risk of colon cancer. Two-thirds of these tumors occur in the proximal colon. Women with HNPCC have a 20-60% lifetime risk of endometrial cancer. In HNPCC, the gastric cancer is usually intestinal-type adenocarcinoma. The ovarian cancer in HNPCC may be diagnosed before age 40. Other HNPCC-related cancers have characteristic features: the urinary tract cancers are transitional carcinoma of the ureter and renal pelvis; the small bowel cancer is most common in the duodenum and jejunum; and the most common type of brain tumor is glioblastoma.

The diagnosis of HNPCC may be made on the basis of the Amsterdam clinical criteria (listed above) or on the basis of molecular genetic testing for mutations in a mismatch repair gene (MLH1, MSH2, MSH6 or PMS2). Mutations in MLH1 and MSH2 account for 90% of HNPCC. Mutations in MSH6 and PMS2 account for the rest.

HNPCC is inherited in an autosomal dominant manner. Each child of an individual with HNPCC has a 50% chance of inheriting the mutation. Most people diagnosed with HNPCC have inherited the condition from a parent. However, not all individuals with an HNPCC gene mutation have a parent who had cancer. Prenatal diagnosis for pregnancies at increased risk for HNPCC is possible.

In tumors that are microsatellite instable it is often found that the DNA mismatch repair proteins that are encoded by the *MLH1* or *MSH2* genes are inactivated. In case of microsatellite instable hereditary non-polyposis colorectal cancers germline mutation in MLH1 and MSH2 and somatic loss of function of the normal allele has been found to be associated with the disease.

For most sporadic MSI tumors epigenetic hypermethylation of the MLH1 promoter can be found to be associated with the cancer (Cunningham JM, Christensen ER, Tester DJ, et al., Cancer Res 1998;58(15):3455-60., Kane MF, Loda M, Gaida GM, et al., Cancer Res 1997;57(5):808-11., Herman JG, Umar A, Polyak K, et al., Proc Natl Acad Sci U S A 1998;95(12):6870-5., Kuismanen SA, Holmberg MT, Salovaara R, de la Chapelle A, Peltomaki P., Am J Pathol 2000;156(5):1773-9).

Forms of cancer

Cancer leads to a change in the expression of one or more genes. The methods according to the invention may be used for classifying cancer according to the microsatellite status and/or the hereditary or sporadic nature of the cancer. Thus, the cancer may be any malignant condition in which genomic instability is involved in the development of cancer, such as cancers related to hereditary non-polyposis colorectal cancer, such as endometrial cancer, gastric cancer, small bowel cancer, ovarian cancer, kidney cancer, pelvic renal cancer or tumors of the nervous system, such as glioblastoma.

One particular form of cancer according to the present invention is that of the colon/rectum.

The cancer may be of any tumor type, such as an adenocarcinoma, a carcinoma, a teratoma, a sarcoma, and/or a lymphoma.

In relation to the gastro-intestinal tract, the biological condition may also be colitis ulcerosa, Mb. Crohn, diverticulitis, adenomas.

Colorectal tumors

The data presented herein relates to colorectal tumors and therefore the description has focused on the gene expression level as one manner of identifying genes involved in the prediction of survival in cancer tissue. The malignant progression of cancer of colon or rectum may be described using Dukes stages where normal mucosa may progress to Dukes A superficial tumors to Dukes B, slightly invasive tumors, to Dukes C that have spread to lymphnodes and finally to Dukes D that have metastasized to other organs.

The grade of a tumor can also be expressed on a scale of I-IV. The grade reflects the cytological appearance of the cells. Grade I cells are almost normal, whereas grade II cells deviate slightly from normal. Grade III appear clearly abnormal, whereas grade IV cells are highly abnormal.

The phrase colon cancer is in this application meant to be equivalent to the phrase colorectal cancer. Colon cancers may be located in the right side of the colon, the left side of the colon, the transverse part of the colon and/or in the rectum.

Samples

The samples according to the present invention may be any cancer tissue.

The sample may be in a form suitable to allow analysis by the skilled artisan, such as a biopsy of the tissue, or a superficial sample scraped from the tissue. In one embodiment of the invention it is preferred that the sample is from a resected colon cancer tumor. In another embodiment the sample may be prepared by forming a suspension of cells made from the tissue. The sample may, however, also be an extract obtained from the tissue or obtained from a cell suspension made from the tissue. The sample may be fresh or frozen, or treated with chemicals.

Expression pattern

Expression of one gene or more genes in a sample forms a pattern that is characteristic of the state of the cell. In a sample from an individual having contracted cancer a plurality of gene expression products are present. By expression pattern is meant

the presence of a combination of a number of expression products and/or the amount of expression products specific for a given biological condition, such as cancer. The pattern is produced by determining the expression products of selected genes that together reveals a pattern that is indicative of the biological condition. Thus, a selection of the genes that carry information about a specific condition is developed. Selection of the genes is achieved by analyzing large numbers of genes and their expression products to find the genes that will enable the desired differentiation between various conditions, such as microsatellite status (MSS or MSI) and/or prognostic marker, such as for example the sporadic or hereditary nature of a given cancer sample. The criteria for selection of the best genes for the pattern to be indicative of given biological conditions include confidence levels i.e. how accurate are the selected genes forming an expression pattern in giving correct information of the biological condition. Thus, in one aspect of the present invention a specific pattern of gene expression profiles can be used to determine the microsatellite status in the sample. In a second aspect of the present invention the microsatellite status is determined and a specific pattern of the presence of a plurality of gene expression products and/or amount wherefrom a prognostic marker is determined.

Determination of the microsatellite status employing gene expression patterns

One aspect of the invention specifically relates to a method for determining the microsatellite status in a sample of an individual having contracted cancer based on determination of the expression pattern of at least two genes, such as at least three genes, such as at least four genes, such as at least 5 genes, such as at least 6 genes, such as at least 7 genes, such as at least 8 genes, such as at least 9 genes, such as at least 10 genes, such as at least 15 genes, such as at least 20 genes, such as at least 30 genes, such as at least 40 genes, such as at least 50 genes, such as at least 60 genes, such as at least 70 genes, such as at least 80 genes, such as at least 90 genes, such as at least 126 genes selected from the group of genes listed in Table 1 below

Table 1

Gene name	Ref seq	Gene symbol	SEQ ID NO.:
chemokine (C-C motif) ligand 5	NM_002985	CCL5	1
tryptophanyl-tRNA synthetase	NM_004184	WARS	2
proteasome (prosome, macropain) activator subunit 1 (PA28 alpha)	NM_006263	PSME1	3
bone marrow stromal cell antigen 2	NM_004335	BST2	4
ubiquitin-conjugating enzyme E2L 6	NM_004223	UBE2L6	5

A kinase (PRKA) anchor protein 1	NM_003488	AKAP1	6
proteasome (prosome, macropain) activator subunit 2 (PA28 beta)	NM_002818	PSME2	7
carcinoembryonic antigen-related cell adhesion molecule 5	NM_004363	CEACAM5	8
FERM, RhoGEF (ARHGEF) and pleckstrin domain protein 1 (chondrocyte-derived)	NM_005766	FARP1	9
myosin X	NM_012334	MYO10	10
heterogeneous nuclear ribonucleoprotein L	NM_001533	HNRPL	11
autocrine motility factor receptor	NM_001144	AMFR	12
dimethylarginine dimethylaminohydrolase 2	NM_013974	DDAH2	13
tumor necrosis factor, alpha-induced protein 2	NM_006291	TNFAIP2	14
mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli)	NM_000249	MLH1	15
thymidylate synthetase	NM_001071	TYMS	16
intercellular adhesion molecule 1 (CD54), human rhinovirus receptor	NM_000201	ICAM1	17
general transcription factor IIA, 2, 12kDa	NM_004492	GTF2A2	18
Rho-associated, coiled-coil containing protein kinase 2	NM_004850	ROCK2	19
ATP binding protein associated with cell differentiation	NM_005783	TXNDC9	20
NCK adaptor protein 2	NM_003581	NCK2	21
phytanoyl-CoA hydroxylase (Refsum disease)	NM_006214	PHYH	22
metastais-associated gene family, member 2	NM_004739	MTA2	23
amiloride binding protein 1 (amine oxidase (copper-containing))	NM_001091	ABP1	24
biliverdin reductase A	NM_000712	BLVRA	25
phospholipase C, beta 4	NM_000933	PLCB4	26
chemokine (C-X-C motif) ligand 9	NM_002416	CXCL9	27
purine-rich element binding protein A	NM_005859	PURA	28
quinolinate phosphoribosyltransferase (nicotinate-nucleotide pyrophosphorylase (carboxylating))	NM_014298	QPRT	29
retinoic acid receptor responder (tazarotene induced) 3	NM_004585	RARRES3	30
chemokine (C-C motif) ligand 4	NM_002984	CCL4	31
forkhead box O3A	NM_001455	FOXO3A	32
interferon, alpha-inducible protein (clone IFI-6-16)	NM_002038 NM_022873	G1P3	34 123
chemokine (C-X-C motif) ligand 10	NM_001565	CXCL10	35
metallothionein 1G	NM_005950 NM_005950	MT1G	36
tumor necrosis factor receptor superfamily, member 6	NM_000043 NM_152877 NM_152876 NM_152875 NM_152872 NM_152873 NM_152871 NM_152874	TNFRSF6	37 133 132 134 130 33 129 131
endothelial cell growth factor 1 (platelet-derived)	NM_001953	ECGF1	38
SCO cytochrome oxidase deficient homolog 2 (yeast)	NM_005138	SCO2	39
chemokine (C-X-C motif) ligand 13 (B-cell chemoattractant)	NM_006419	CXCL13	40

Granulysin	NM_006433	GNLY	41
CD2 antigen (p50), sheep red blood cell receptor	NM_001767	CD2	42
splicing factor, arginine/serine-rich 6	NM_006275	SFRS6	43
teratocarcinoma-derived growth factor 1	NM_003212	TDGF1	44
metallothionein 1H	NM_005951	MT1H	45
cytochrome P450, family 2, subfamily B, polypeptide 6	NM_000767	CYP2B6	46
tumor necrosis factor (ligand) superfamily, member 9	NM_003811	TNFSF9	47
RNA binding motif protein 12	NM_006047 NM_006047	RBM12	48
heat shock 105kDa/110kDa protein 1	NM_006644	HSPH1	49
staufen, RNA binding protein (Drosophila)	NM_004602 NM_017452 NM_017453	STAU	50 125 126
lymphocyte antigen 6 complex, locus G6D	NM_021246	LY6G6D	51
calcium binding protein P22	NM_007236	CHP	52
CDC14 cell division cycle 14 homolog B (S. cerevisiae)	NM_003671 NM_033331	CDC14B	53 115
epiplakin 1	XM_372063	EPPK1	54
metallothionein 1X	NM_005952	MT1X	55
transforming growth factor, beta receptor II (70/80kDa)	NM_003242	TGFBR2	56
protein kinase C binding protein 1	NM_012408 NM_183047	PRKCBP1	57 124
transmembrane 4 superfamily member 6	NM_003270	TM4SF6	58
pleckstrin homology domain containing, family B (evectins) member 1	NM_021200	PLEKHB1	59
apolipoprotein L, 1	NM_003661 NM_145343	APOL1	60 120
indoleamine-pyrrole 2,3 dioxygenase	NM_002164	INDO	61
forkhead box A2	NM_021784	FOXA2	62
granzyme H (cathepsin G-like 2, protein h-CCPX)	NM_033423	GZMH	63
baculoviral IAP repeat-containing 3	NM_001165	BIRC3	64
Homo sapiens metallothionein 1H-like protein		AF333388 (Hs 382039)	135
KIAA0182 protein	NM_014615	KIAA0182	117
G protein-coupled receptor 56	NM_005682	GPR56	65

	NM_201524		116
metallothionein 2A	NM_005953	MT2A	66
F-box only protein 21	NM_015002	FBXO21	67
	NM_012156,	EPB41L1	68
erythrocyte membrane protein band 4.1-like 1	NM_012156		
hypothetical protein MGC21416	NM_173834	MGC21416	69
protein O-fucosyltransferase 1	NM_015352, NM_015352	POFUT1	70
metallothionein 1E (functional)	NM_175617	MT1E	71
troponin T1, skeletal, slow	NM_003283	TNNT1	72
chimerin (chimaerin) 2	NM_004067	CHN2	73
heterogeneous nuclear ribonucleoprotein H1 (H)	NM_005520	HNRPH1	74
ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit, isoform 1, cardiac muscle	NM_004046	ATP5A1	75
eukaryotic translation initiation factor 5A	NM_001970	EIF5A	76
perforin 1 (pore forming protein)	NM_005041	PRF1	77
OGT(O-Glc-NAc transferase)-interacting protein 106 KDa	NM_014965	OIP106	78
DEAD (Asp-Glu-Ala-Asp) box polypeptide 27	NM_017895	DDX27	79
vacuolar protein sorting 35 (yeast)	NM_018206	VPS35	80
tripartite motif-containing 44	NM_017583	TRIM44	81
transmembrane, prostate androgen induced RNA	NM_020182	TMEPAI	82
	NM_199169		127
	NM_199170		128
dynein, cytoplasmic, light polypeptide 2A	NM_014183	DNCL2A	83
	NM_177953		122
leucine aminopeptidase 3	NM_015907	LAP3	84
chromosome 20 open reading frame 35	NM_018478	C20orf35	85
	NM_033542		118
solute carrier family 38, member 1	NM_030674	SLC38A1	86
CGI-85 protein	NM_016028	CGI-85	87
death associated transcription factor 1	NM_022105,	DATF1	88
	NM_080796		121
hepatocellular carcinoma-associated antigen 112	NM_018487	HCA112	89
sestrin 1	NM_014454	SESN1	90
hypothetical protein FLJ20315	NM_017763	FLJ20315	91
hypothetical protein FLJ20647	NM_017918	FLJ20647	92
membrane protein expressed in epithelial-like lung adenocarcinoma	NM_024792	CT120	93
DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide	NM_014314	RIG-I	94
keratin 23 (histone deacetylase inducible)	NM_015515,	KRT23	95
UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-	NM_007210	GALNT6	96

acetylgalactosaminyltransferase 6 (GalNAc-T6)			
aryl hydrocarbon receptor nuclear translocator-like 2	NM_020183	ARNTL2	97
apobec-1 complementation factor	NM_014576,	ACF	98
	NM_138932		119
hypothetical protein FLJ20232	NM_019008	FLJ20232	99
apolipoprotein L, 2	NM_030882,	APOL2	100
	NM_145343		120
mitochondrial solute carrier protein	NM_016612	MSCP	101
hypothetical protein FLJ20618	NM_017903	FLJ20618	102
	NM_003011.		103
SET translocation (myeloid leukaemia-associated)	1	SET	
	Xm_030577.		104
ATPase, class II, type 9a	9	ATP9a	

One embodiment of the invention concerning the determination of microsatellite status is based on the expression pattern of at least 2 genes, such as at least 3 genes, such as at least 4 genes, such as at least 5 genes, such as at least 6 genes, such as at least 7 genes, such as at least 8 genes, such as at least 9 genes, such as at least 10 genes, such as at least 15 genes, such as at least 20 genes, such as at least 25 genes selected from the group of genes listed in Table 2.

Table 2

Gene name	Ref seq	Gene symbol	SEQ ID NO.:
chemokine (C-C motif) ligand 5	NM_002985	CCL5	1
tryptophanyl-tRNA synthetase	NM_004184	WARS	2
proteasome (prosome, macropain) activator subunit 1 (PA28 alpha)	NM_006263	PSME1	3
bone marrow stromal cell antigen 2	NM_004335	BST2	4
ubiquitin-conjugating enzyme E2L 6	NM_004223	UBE2L6	5
A kinase (PRKA) anchor protein 1	NM_003488	AKAP1	6
proteasome (prosome, macropain) activator subunit 2 (PA28 beta)	NM_002818	PSME2	7
carcinoembryonic antigen-related cell adhesion molecule 5	NM_004363	CEACAM5	8
FERM, RhoGEF (ARHGEF) and pleckstrin domain protein 1 (chondrocyte-derived)	NM_005766	FARP1	9
myosin X	NM_012334	MYO10	10
heterogeneous nuclear ribonucleoprotein L	NM_001533	HNRPL	11
autocrine motility factor receptor	NM_001144	AMFR	12
dimethylarginine dimethylaminohydrolase 2	NM_013974	DDAH2	13
tumor necrosis factor, alpha-induced protein 2	NM_006291	TNFAIP2	14
mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli)	NM_000249	MLH1	15
thymidylate synthetase	NM_001071	TYMS	16

intercellular adhesion molecule 1 (CD54), human rhinovirus receptor	NM_000201	ICAM1	17
general transcription factor IIA, 2, 12kDa	NM_004492	GTF2A2	18
Rho-associated, coiled-coil containing protein kinase 2	NM_004850	ROCK2	19
ATP binding protein associated with cell differentiation	NM_005783	APACD	20
metastais-associated gene family, member 2	NM_004739	MTA2	23
chemokine (C-X-C motif) ligand 10	NM_001565	CXCL10	35
splicing factor, arginine/serine-rich 6	NM_006275	SFRS6	43
protein kinase C binding protein 1	NM_012408 NM_183047	PRKCBP1	57 124
hepatocellular carcinoma-associated antigen 112	NM_018487	HCA112	89
hypothetical protein FLJ20618	NM_017903	FLJ20618	102
SET translocation (myeloid leukaemia-associated)	NM_003011.1	SET	103
ATPase, class II, type 9a	Xm_030577.9	ATP9a	104

or from

Table 3

Gene name	Ref seq	Gene symbol	SEQ ID NO.:
heterogeneous nuclear ribonucleoprotein L	NM_001533	HNRPL	11
NCK adaptor protein 2	NM_003581	NCK2	21
phytanoyl-CoA hydroxylase (Refsum disease)	NM_006214	PHYH	22
metastais-associated gene family, member 2	NM_004739	MTA2	23
amiloride binding protein 1 (amine oxidase (copper-containing))	NM_001091	ABP1	24
biliverdin reductase A	NM_000712	BLVRA	25
phospholipase C, beta 4	NM_000933	PLCB4	26
chemokine (C-X-C motif) ligand 9	NM_002416	CXCL9	27
purine-rich element binding protein A	NM_005859	PURA	28
quinolinate phosphoribosyltransferase (nicotinate-nucleotide pyrophosphorylase (carboxylating))	NM_014298	QPRT	29
retinoic acid receptor responder (tazarotene induced) 3	NM_004585	RARRES3	30
chemokine (C-C motif) ligand 4	NM_002984	CCL4	31
forkhead box O3A	NM_001455	FOXO3A	32
metallothionein 1X	NM_005952	MT1X	55
interferon, alpha-inducible protein (clone IFI-6-16)	NM_002038 NM_022873	G1P3	34 123
chemokine (C-X-C motif) ligand 10	NM_001565	CXCL10	35
metallothionein 1G	NM_005950, NM_005950	MT1G	36

tumor necrosis factor receptor superfamily, member 6	NM_000043 NM_152877 NM_152876 NM_152875 NM_152872 NM_152873 NM_152871 NM_152874	TNFRSF6	37 133 132 134 130 33 129 131
endothelial cell growth factor 1 (platelet-derived)	NM_001953	ECGF1	38
SCO cytochrome oxidase deficient homolog 2 (yeast)	NM_005138	SCO2	39
chemokine (C-X-C motif) ligand 13 (B-cell chemoattractant)	NM_006419	CXCL13	40
Granulysin	NM_006433	GNLY	41
splicing factor, arginine/serine-rich 6	NM_006275	SFRS6	43
protein kinase C binding protein 1	NM_012408 NM_183047	PRKCBP1	57 124
hepatocellular carcinoma-associated antigen 112	NM_018487	HCA112	89
hypothetical protein FLJ20618	NM_017903	FLJ20618	102
SET translocation (myeloid leukaemia-associated)	NM_003011.1	SET	103
ATPase, class II, type 9a	Xm_030577.9	ATP9a	104

or from

Table 4

Gene name	Ref seq	Gene symbol	SEQ ID NO.:
heterogeneous nuclear ribonucleoprotein L	NM_001533	HNRPL	11
metastais-associated gene family, member 2	NM_004739	MTA2	23
chemokine (C-X-C motif) ligand 10	NM_001565	CXCL10	35
CD2 antigen (p50), sheep red blood cell receptor	NM_001767	CD2	42
splicing factor, arginine/serine-rich 6	NM_006275	SFRS6	43
teratocarcinoma-derived growth factor 1	NM_003212	TDGF1	44
metallothionein 1H	NM_005951	MT1H	45
cytochrome P450, family 2, subfamily B, polypeptide 6	NM_000767	CYP2B6	46
tumor necrosis factor (ligand) superfamily, member 9	NM_003811	TNFSF9	47
RNA binding motif protein 12	NM_006047, NM_006047	RBM12	48
heat shock 105kDa/110kDa protein 1	NM_006644	HSPH1	49
staufen, RNA binding protein (Drosophila)	NM_004602 NM_017452 NM_017453	STAU	50 125 126
lymphocyte antigen 6 complex, locus G6D	NM_021246	LY6G6D	51
calcium binding protein P22	NM_007236	CHP	52

CDC14 cell division cycle 14 homolog B (S. cerevisiae)	NM_003671 NM_033331	CDC14B	53 115
epiplakin 1	XM_372063	EPPK1	54
metallothionein 1X	NM_005952	MT1X	55
transforming growth factor, beta receptor II (70/80kDa)	NM_003242	TGFBR2	56
protein kinase C binding protein 1	NM_012408 NM_183047	PRKCBP1	57 129
transmembrane 4 superfamily member 6	NM_003270	TM4SF6	58
pleckstrin homology domain containing, family B (evectins) member 1	NM_021200	PLEKHB1	59
apolipoprotein L, 1	NM_003661 NM_145343	APOL1	60 125
indoleamine-pyrrole 2,3 dioxygenase	NM_002164	INDO	61
	NM_021784	FOXA2	62
forkhead box A2	NM_021784		
hepatocellular carcinoma-associated antigen 112	NM_018487	HCA112	89
mitochondrial solute carrier protein	NM_016612 NM_016612	MSCP	101
hypothetical protein FLJ20618	NM_017903	FLJ20618	102
SET translocation (myeloid leukaemia-associated)	NM_003011.1	SET	103
ATPase, class II, type 9a	Xm_030577.9	ATP9a	104

or from

Table 5

Gene name	Ref seq	Gene symbol	SEQ ID NO.:
heterogeneous nuclear ribonucleoprotein L	NM_001533	HNRPL	11
metastals-associated gene family, member 2	NM_004739	MTA2	23
chemokine (C-X-C motif) ligand 10	NM_001565	CXCL10	35
splicing factor, arginine/serine-rich 6	NM_006275	SFRS6	43
protein kinase C binding protein 1	NM_012408 NM_183047	PRKCBP1	57 124
granzyme H (cathepsin G-like 2, protein h-CCPX)	NM_033423	GZMH	63
	NM_001165	BIRC3	64
baculoviral IAP repeat-containing 3	NM_001165		
Homo sapiens metallothionein 1H-like protein		AF333388 (Hs 382039)	135
KIAA0182 protein	NM_014615	KIAA0182	117
	NM_005682	GPR56	65
G protein-coupled receptor 56	NM_301524		116

metallothionein 2A	NM_005953	MT2A	66
F-box only protein 21	NM_015002	FBXO21	67
erythrocyte membrane protein band 4.1-like 1	NM_012156	EPB41L1	68
hypothetical protein MGC21416	NM_173834	MGC21416	69
protein O-fucosyltransferase 1	NM_015352	POFUT1	70
metallothionein 1E (functional)	NM_175617	MT1E	71
troponin T1, skeletal, slow	NM_003283	TNNT1	72
chimerin (chimaerin) 2	NM_004067	CHN2	73
heterogeneous nuclear ribonucleoprotein H1 (H)	NM_005520	HNRPH1	74
ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit, isoform 1, cardiac muscle	NM_004046	ATP5A1	75
eukaryotic translation initiation factor 5A	NM_001970	EIF5A	76
perforin 1 (pore forming protein)	NM_005041	PRF1	77
OGT(O-Glc-NAc transferase)-interacting protein 106 KDa	NM_014965	OIP106	78
DEAD (Asp-Glu-Ala-Asp) box polypeptide 27	NM_017895	DDX27	79
hepatocellular carcinoma-associated antigen 112	NM_018487	HCA112	89
hypothetical protein FLJ20232	NM_019008	FLJ20232	99
apolipoprotein L, 2	NM_030882, NM_145343	APOL2	100 120
hypothetical protein FLJ20618	NM_017903	FLJ20618	102
SET translocation (myeloid leukaemia-associated)	NM_003011.1	SET	103
ATPase, class II, type 9a	Xm_030577.9	ATP9a	104

or from

Table 6

Gene name	Ref seq	Gene sym-bol	SEQ ID NO.:
heterogeneous nuclear ribonucleoprotein L	NM_001533	HNRPL	11
metastais-associated gene family, member 2	NM_004739	MTA2	23
chemokine (C-X-C motif) ligand 10	NM_001565	CXCL10	35
metallothionein 1G	NM_005950	MT1G	36
splicing factor, arginine/serine-rich 6	NM_006275	SFRS6	43
protein kinase C binding protein 1	NM_012408 NM_183047	PRKCBP1	57 129
vacuolar protein sorting 35 (yeast)	NM_018206	VPS35	80
tripartite motif-containing 44	NM_017583	TRIM44	81

	NM_020182 NM_199169	TMEPAI	82 127
transmembrane, prostate androgen induced RNA	NM_199170		128
dynein, cytoplasmic, light polypeptide 2A	NM_014183 NM_177953	DNCL2A	83 122
leucine aminopeptidase 3	NM_015907	LAP3	84
chromosome 20 open reading frame 35	NM_018478 NM_033542	C20orf35	85 118
solute carrier family 38, member 1	NM_030674	SLC38A1	86
CGI-85 protein	NM_016028	CGI-85	87
death associated transcription factor 1	NM_022105, NM_080796	DATF1	88 121
hepatocellular carcinoma-associated antigen 112	NM_018487	HCA112	89
sestrin 1	NM_014454	SESN1	90
hypothetical protein FLJ20315	NM_017763	FLJ20315	91
hypothetical protein FLJ20647	NM_017918	FLJ20647	92
membrane protein expressed in epithelial-like lung adenocarcinoma	NM_024792	CT120	93
DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide	NM_014314	RIG-I	94
keratin 23 (histone deacetylase inducible)	NM_015515	KRT23	95
UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 6 (GalNAc-T6)	NM_007210	GALNT6	96
aryl hydrocarbon receptor nuclear translocator-like 2	NM_020183	ARNTL2	97
apobec-1 complementation factor	NM_014576 NM_138932	ACF	98 119
hypothetical protein FLJ20618	NM_017903	FLJ20618	102
SET translocation (myeloid leukaemia-associated)	NM_003011.1	SET	103
ATPase, class II, type 9a	Xm_030577.9	ATP9a	104

Another embodiment of the invention concerning the determination of microsatellite status is based on the expression pattern of at least 2 genes, such as at least 3 genes, such as at least 4 genes, such as at least 5 genes, such as at least 6 genes, such as at least 7 genes, such as at least 8 genes, such as at least 9 genes selected from the group of genes listed in Table 7 below.

RNA purification Colon specimens were obtained fresh from surgery and were immediately snap frozen in liquid nitrogen either as was, in OCD-compound or in a SDS/guadinium thiocyanate solution. Total RNA was isolated using RNAzol (WAK-Chemie Medical) or spin column technology (Sigma) following the manufactures' instructions.

Gene expression analysis These procedures were performed at described in detail elsewhere (Dyrskødt et al). Briefly, ten μg of total RNA was used as starting material for the target preparation as described. First and second strand cDNA synthesis was performed using the SuperScript II System (Invitrogen) according to the manufacturers' instructions except using an oligo-dT primer containing a T7 RNA polymerase promoter site. Labelled aRNA was prepared using the BioArray High Yield RNA Transcript Labelling Kit (Enzo) using Biotin labelled CTP and UTP (Enzo) in the reaction together with unlabeled NTP´s. Unincorporated nucleotides were removed using RNeasy columns (Qiagen). Fifteen μg of cRNA was fragmented, loading onto the Affymetrix HG_U133A probe array cartridge and hybridized for 16h. The arrays were washed and stained in the Affymetrix Fluidics Station and scanned using a confocal laser-scanning microscope (Hewlett Packard GeneArray Scanner G2500A). The readings from the quantitative scanning were analyzed by the Affymetrix Gene Expression Analysis Software (MAS 5.0) and normalized using RMA (robust multi array normalisation, Irizarry et al. 2002) in the statistical application R. Redundant probesets (as defined form Unigene build 168) with high correlation (>0.5) over all samples were removed, which reduced the dataset to approximately 14.400 probesets. This dataset was used a source for all further calculations in this manuscript.

Unsupervised agglomerative hierarchical clustering

For hierarchical cluster analysis 1239 genes with a variation across all samples greater than 0.5 were median-centred to a magnitude of 1. Samples and genes were then clustered using average linkage clustering with a modified Person correlation as similarity metric (Eisen et al., PNAS 95: 14863-14868, 1998). The cluster dendrogram was visualized with TreeView (Eisen).

Group testing

We make a statistical test where the p-value is evaluated through permutations. For each group and gene we calculate the average and the sum of squared deviations from the average. We then sum these over the genes and the groups:

This expression is calculated for joining DK with SF and MSI with MSS such that we end up with two groups. The sum of squared deviations is denoted S_2. As a test statistic we use S_1/S_2. A small value indicates that there is a real reduction in the deviations when going from 2 to 4 groups and thus the groups have a real significance. To judge if a value is significantly small we use permutations. For each of the four groups left when joining DK and SF we randomly allocate the members to a pseudo DK and pseudo SF in such a way that the number of members in each group are as in the original data.

$$S_1 = \sum_{groups} \sum_{genes} \left(X_{ij} - \bar{X}_{gr(i)j}\right)^2$$

To get an understanding of this separation we performed a test to see if this is caused by few genes or if many genes are involved. For this test we calculated $S_1 = \sum_{genes} S_1(gene)$ and similarly with $S_2 = \sum_{genes} S_2(gene)$. For each gene j we used the test statistic $S_1(j)/S_2(j)$ (Table 3).

Multidimentional scaling

We carried out multidimentional scaling on median-centered and normalized data using CMD—scale in the statistical application R and visualized in a two-dimentional plot.

Microsatellite status classifier

The readings from the quantitative scanning were analyzed by the Affymetrix Gene Expression Analysis Software (MAS 5.0) and normalized using RMA (robust multi array normalisation, Irizarry et al. 2002) in the statistical application R. Redundant probesets (as defined form Unigene build 168) with high correlation (>0.5) over all samples were removed, which reduced the dataset to approximately 14.400 probesets.

The microsatellite instability status classifier was based on a dataset of 4.266 genes. These genes result from the removal of genes with a variance over all tumor sam-

ples smaller than 0.2 and genes that separate Danish from Finnish samples with a t-value numerically greater than 2. We used a normal distribution with the mean dependent on the gene and the group (MSI, MSS). For each gene, we calculated the variation between the groups and the variation within the groups to select genes with a high ratio between these. To classify a sample, we calculated the sum over the genes of the squared distance from the sample value to the group mean, standardized by the variance and assigned the sample to the nearest group. The sample to be classified was excluded when calculating group means and variances.

Estimation of classifier stability

We validated the performance of the classifier by permutation. One hundred datasets consisting of 30 MSS samples and 25 MSI samples were randomly chosen by permutation for training of the classifier with the remaining samples in each case being assign to a testset. Averages over the 100 data sets of the number of errors in the cross-validation of the training set and in the test set were used as a measure of the precision of the classifier.

Real-time PCR (RT-PCR). The procedures were as described (Birkenkamp-Demtroder) except that we used short LNA (Locked Nucleic Acid) enhanced probes from a Human Probe Library (Exiqon™). In short, cDNA was synthesized from single samples some of which were previously analyzed on GeneChips. Reverse transcription was performed using Superscript II RT (Invitrogen). Real-time PCR analysis was performed on selected genes using the primers (DNA Technology) and probes (Exiqon, DK) described in figure legend X. All samples were normalized to GAPDH as described previously (Birkenkamp-Demtroder et. al. Cancer Res., *62*: 4352-4363, 2002).

Rebuilding of Classifier based on Real-Time PCR

The 79 tumors samples that were not analysed by real-time PCR were transformed into log ratios using one of the tumor samples as reference and used for training of the classifier. Then 23 samples of which 18 were also analyzed on arrays were equally transformed into log ratios using the same tumor sample as above as reference and tested. The idea behind this translation is that we expect the normalized PCR values to be proportional to the normalized array values, and on a log scale this becomes an additive difference. The difference is gene specific and is therefore

estimated for each gene separately. The variation obtained from the microarray data, and used in the classifier, can be used directly on the PCR platform.

Results

Hierarchical clustering

The clinical specimens used in this study were collected in two different countries from 14 different clinics in the period 1994 to 2001. The samples were selected to keep a balanced representation of microsatellite instable (MSI) and microsatellite stable (MSS) tumors from both the right- and left-sided colon. The MSI class was represented both by sporadic MSI and hereditary MSI (HNPCC) tumors. Only Dukes' B and Dukes' C tumor samples were included were selected (table 19). Before any attempt to divide a diverse sample collection into distinct classes analyzed the data for systematic bias that may have been introduces during the experimental procedures. A fast and easy way to discover both true distinct classes as well as systematic biases in the data is to perform a hierarchical clustering.

The phylogenetic tree resulting from hierarchical clustering on 1239 genes (Fig. 6) reveals that the main separating factor is microsatellite status. On the upper trunk we find two clusters represented mainly by normal biopsies (14/21) and MSS tumors (18/25), respectively. The lower trunk is divided into a MSI cluster (30/36) and a second MSS cluster (MSS2-cluster) (34/37). A closer inspection of the two MSS clusters unveil that one is dominated by Danish samples (19/25) and one by Finnish samples (26/37 check). Also, it is worth to notice that the MSI cluster contains a vast majority of Finnish samples (32/36) and that the sporadic MSI samples are interspersed among the hereditary samples. The normal biopsies cluster tight together with a slight tendency to separation according to origin. Tree normal samples cluster within the MSI cluster indicating that resection of these samples may have been to close to the tumor lesion.

Inspection of the gene cluster dendrogram shows that the two groups of MSS tumors are mainly separated by a large cluster of genes being upregulated in the Danish samples (data not shown) indicating that a systematic difference between Danish and Finnish samples.

Significance of observed groups

Based on these observations, we performed a series of test to evaluate if the observed separation of tumors into MSS and MSI as well as DK and SF are significant. For these tests the tumor samples were grouped into four virtual tumor-groups labelled, i.e. Danish MSI (MSI-DK), Danish MSS (MSS-DK), Finnish MSI (MSI-SF) and Finnish MSS (MSS-SF). Based on 5082 genes with a variance above 0.2, we tested if all four groups are significant or if some of the groups can be joined. We considered the two possibilities of joining DK and SF, and of joining MSI and MSS and made a statistical test where the p-value is evaluated through permutations. In 100 permutations of each group combination our test value S1/S2 is considerably smaller than in all permutation (Table 20) demonstrating a very clear separation between DK and SF and between MSI and MSS.

Table 20

Permutation test of groups

Pseudo group	S1/S2 from data	Smaller values in 100 permutations	Minimum in 100 permutations
DK-SF	0.9072795	0	0.962269
I-S	0.9166195	0	0.9583325

Such a clear distinction between groups may rely on a few highly separating genes or a general difference in the gene expression profile including many genes. For both the DK-SF and MSI-MSS the effect are caused by many genes even at very criteria, i.e. low test statistic $S_1(j)/S_2(j)$ values (Table 21).

Table 21

Permutation test of genes

		$S_1(j)/S_2(j)$			
Pseudo group		< 0.6	< 0.7	< 0.8	< 0.9
DK-SF	number of genes	36	136	522	1785
	max in 100 permutations	0	0	2	225
MSI-MSS	number of genes	17	103	399	1507
	max in 100 permutations	0	1	8	250

When a property is present that influences a large proportion of the genes this may obscure separation of clinical relevant features in unsupervised clustering. To visual-

ize the effect of such properties, we calculated distances by multidimensional scaling between samples with and without of 816 genes separating DK from SF with a t-value numerically greater than 2 (Fig 7). We see an improved separation of MSI and MSS with Danish and Finnish cases mixed. The MSI-DK samples are not completely separated as they are found both between the MSI-SF and the MSS samples. (These plots are not entirely unsupervised since the groups have been used to remove gene).

Construction of an MSI-MSS classifier

For the construction of a classifier we used the expression profiles from 97 tumors for which no ambiguity had been identified in relation to microsatellite status. The 816 genes separating DK from SF were excluded, as these would be unreliable for MS classification. We built a maximum likelihood classifier in order to select a minimum of genes giving the largest possible separation of the two groups. We tested the performance of the classifier using 1-1000 genes and found that it was stable showing 3-6 errors when using 4 – 400 genes. Of these 106 genes were especially suited for discrimination of MSS from MSI (table 22).

Table 22

AFFYID	*SYMBOL*	*LOCUS LINK*	*OMIM*	*REFSEQ*	*GENENAME*
1405_i_at	CCL5	6352	187011	NM_002985	chemokine (C-C motif) ligand 5
200628_s_at	WARS	7453	191050	NM_004184	tryptophanyl-tRNA synthetase
200814_at	PSME1	5720	600654	NM_006263	proteasome (prosome, macropain) activator subunit 1 (PA28 alpha)
201641_at	BST2	684	600534	NM_004335	bone marrow stromal cell antigen 2
201649_at	UBE2L6	9246	603890	NM_004223	ubiquitin-conjugating enzyme E2L 6
201674_s_at	AKAP1	8165	602449	NM_003488	A kinase (PRKA) anchor protein 1
201762_s_at	PSME2	5721	602161	NM_002818	proteasome (prosome, macropain) activator subunit 2 (PA28 beta)
201884_at	CEACAM5	1048	114890	NM_004363	carcinoembryonic antigen-related cell adhesion molecule 5
201910_at	FARP1	10160	602654	NM_005766	FERM, RhoGEF (ARHGEF) and pleckstrin domain protein 1 (chondrocyte-derived)
201976_s_at	MYO10	4651	601481	NM_012334	myosin X
202072_at	HNRPL	3191	603083	NM_001533	heterogeneous nuclear ribonucleoprotein L
202203_s_at	AMFR	267	603243	NM_001144	autocrine motility factor receptor
202262_x_at	DDAH2	23564	604744	NM_013974	dimethylarginine dimethylaminohydrolase 2
202510_s_at	TNFAIP2	7127	603300	NM_006291	tumor necrosis factor, alpha-induced protein 2
202520_s_at	MLH1	4292	120436	NM_000249	mutL homolog 1, colon cancer, nonpolyposis type 2 (E. coli)
202589_at	TYMS	7298	188350	NM_001071	thymidylate synthetase

202637_s_at	ICAM1	3383	147840	NM_000201	intercellular adhesion molecule 1 (CD54), human rhinovirus receptor
202678_at	GTF2A2	2958	600519	NM_004492	general transcription factor IIA, 2, 12kDa
202762_at	ROCK2	9475	604002	NM_004850	Rho-associated, coiled-coil containing protein kinase 2
203008_x_at	APACD	10190		NM_005783	ATP binding protein associated with cell differentiation
203315_at	NCK2	8440	604930	NM_003581	NCK adaptor protein 2
203335_at	PHYH	5264	602026	NM_006214	phytanoyl-CoA hydroxylase (Refsum disease)
203444_s_at	MTA2	9219	603947	NM_004739	metastais-associated gene family, member 2
203559_s_at	ABP1	26	104610	NM_001091	amiloride binding protein 1 (amine oxidase (copper-containing))
203773_x_at	BLVRA	644	109750	NM_000712	biliverdin reductase A
203896_s_at	PLCB4	5332	600810	NM_000933	phospholipase C, beta 4
203915_at	CXCL9	4283	601704	NM_002416	chemokine (C-X-C motif) ligand 9
204020_at	PURA	5813	600473	NM_005859	purine-rich element binding protein A
204044_at	QPRT	23475	606248	NM_014298	quinolinate phosphoribosyltransferase (nicotinate-nucleotide pyrophosphorylase (carboxylating))
204070_at	RARRES3	5920	605092	NM_004585	retinoic acid receptor responder (tazarotene induced) 3
204103_at	CCL4	6351	182284	NM_002984	chemokine (C-C motif) ligand 4
204131_s_at	FOXO3A	2309	602681	NM_001455	forkhead box O3A
204326_x_at	MT1X	4501	156359	NM_005952	metallothionein 1X
204415_at	G1P3	2537	147572	NM_002038, NM_022873	interferon, alpha-inducible protein (clone IFI-6-16)
204533_at	CXCL10	3627	147310	NM_001565	chemokine (C-X-C motif) ligand 10
204745_x_at	MT1G	4495	156353	NM_005950, NM_005950	metallothionein 1G
204780_s_at	TNFRSF6	355	134637	NM_000043, NM_152877, NM_152876, NM_152875, NM_152872, NM_152873, NM_152871	tumor necrosis factor receptor superfamily, member 6
204858_s_at	ECGF1	1890	131222	NM_001953	endothelial cell growth factor 1 (platelet-derived)
205241_at	SCO2	9997	604272	NM_005138	SCO cytochrome oxidase deficient homolog 2 (yeast)
205242_at	CXCL13	10563	605149	NM_006419	chemokine (C-X-C motif) ligand 13 (B-cell chemoattractant)
205495_s_at	GNLY	10578	188855	NM_006433, NM_006433	granulysin
205831_at	CD2	914	186990	NM_001767	CD2 antigen (p50), sheep red blood cell receptor
206108_s_at	SFRS6	6431	601944	NM_006275	splicing factor, arginine/serine-rich 6
206286_s_at	TDGF1	6997	187395	NM_003212	teratocarcinoma-derived growth factor 1
206461_x_at	MT1H	4496	156354	NM_005951	metallothionein 1H
206754_s_at	CYP2B6	1555	123930	NM_000767	cytochrome P450, family 2, subfamily B, polypeptide 6
206907_at	TNFSF9	8744	606182	NM_003811	tumor necrosis factor (ligand) superfamily, member 9

206918_s_at	RBM12	10137	607179	NM_006047, NM_006047	RNA binding motif protein 12
206976_s_at	HSPH1	10808		NM_006644	heat shock 105kDa/110kDa protein 1
207320_x_at	STAU	6780	601716	NM_004602, NM_004602, NM_017452, NM_017453	staufen, RNA binding protein (Drosophila)
207457_s_at	LY6G6D	58530	606038	NM_021246	lymphocyte antigen 6 complex, locus G6D
207993_s_at	CHP	11261	606988	NM_007236	calcium binding protein P22
208022_s_at	CDC14B	8555	603505	NM_003671, NM_003671, NM_033331	CDC14 cell division cycle 14 homolog B (S. cerevisiae)
208156_x_at	EPPK1	83481			epiplakin 1
208581_x_at	MT1X	4501	156359	NM_005952	metallothionein 1X
208944_at	TGFBR2	7048	190182	NM_003242	transforming growth factor, beta receptor II (70/80kDa)
209048_s_at	PRKCBP1	23613		NM_012408, NM_012408, NM_183047	protein kinase C binding protein 1
209108_at	TM4SF6	7105	300191	NM_003270	transmembrane 4 superfamily member 6
209504_s_at	PLEKHB1	58473	607651	NM_021200	pleckstrin homology domain containing, family B (evectins) member 1
209546_s_at	APOL1	8542	603743	NM_003661, NM_003661, NM_145343	apolipoprotein L, 1
210029_at	INDO	3620	147435	NM_002164	indoleamine-pyrrole 2,3 dioxygenase
210103_s_at	FOXA2	3170	600288	NM_021784, NM_021784	forkhead box A2
210321_at	GZMH	2999	116831	NM_033423	granzyme H (cathepsin G-like 2, protein h-CCPX)
210538_s_at	BIRC3	330	601721	NM_001165, NM_001165	baculoviral IAP repeat-containing 3
211456_x_at	AF333388				
212057_at	KIAA0182	23199		XM_050495	KIAA0182 protein
212070_at	GPR56	9289	604110	NM_005682	G protein-coupled receptor 56
212185_x_at	MT2A	4502	156360	NM_005953	metallothionein 2A
212229_s_at	FBXO21	23014		NM_015002, NM_015002	F-box only protein 21
212336_at	EPB41L1	2036	602879	NM_012156, NM_012156	erythrocyte membrane protein band 4.1-like 1
212341_at	MGC21416	286451		NM_173834	hypothetical protein MGC21416
212349_at	POFUT1	23509	607491	NM_015352, NM_015352	protein O-fucosyltransferase 1
212859_x_at	MT1E	4493	156351	NM_175617	metallothionein 1E (functional)
213201_s_at	TNNT1	7138	191041	NM_003283, NM_003283, XM_352926	troponin T1, skeletal, slow
213385_at	CHN2	1124	602857	NM_004067	chimerin (chimaerin) 2
213470_s_at	HNRPH1	3187	601035	NM_005520	heterogeneous nuclear ribonucleoprotein H1 (H)
213738_s_at	ATP5A1	498	164360	NM_004046	ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit, isoform 1, cardiac muscle
213757_at	EIF5A	1984	600187	NM_001970	eukaryotic translation initiation factor 5A

214617_at	PRF1	5551	170280	NM_005041	perforin 1 (pore forming protein)
214924_s_at	OIP106	22906	608112	NM_014965	OGT(O-Glc-NAc transferase)-interacting protein 106 KDa
215693_x_at	DDX27	55661		NM_017895	DEAD (Asp-Glu-Ala-Asp) box polypeptide 27
215780_s_at	Hs.382039				
216336_x_at	AL031602				
217727_x_at	VPS35	55737	606931	NM_018206	vacuolar protein sorting 35 (yeast)
217759_at	TRIM44	54765		NM_017583	tripartite motif-containing 44
217875_s_at	TMEPAI	56937	606564	NM_020182, NM_020182, NM_199169, NM_199170	transmembrane, prostate androgen induced RNA
217917_s_at	DNCL2A	83658	607167	NM_014183, NM_014183, NM_177953	dynein, cytoplasmic, light polypeptide 2A
217933_s_at	LAP3	51056	170250	NM_015907	leucine aminopeptidase 3
218094_s_at	C20orf35	55861		NM_018478, NM_018478	chromosome 20 open reading frame 35
218237_s_at	SLC38A1	81539		NM_030674	solute carrier family 38, member 1
218242_s_at	CGI-85	51111		NM_016028, NM_016028	CGI-85 protein
218325_s_at	DATF1	11083	604140	NM_022105, NM_022105, NM_080796	death associated transcription factor 1
218345_at	HCA112	55365		NM_018487	hepatocellular carcinoma-associated antigen 112
218346_s_at	SESN1	27244	606103	NM_014454	sestrin 1
218704_at	FLJ20315	54894		NM_017763	hypothetical protein FLJ20315
218802_at	FLJ20647	55013		NM_017918	hypothetical protein FLJ20647
218898_at	CT120	79850		NM_024792	membrane protein expressed in epithelial-like lung adenocarcinoma
218943_s_at	RIG-I	23586		NM_014314	DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide
218963_s_at	KRT23	25984	606194	NM_015515, NM_015515	keratin 23 (histone deacetylase inducible)
219956_at	GALNT6	11226	605148	NM_007210	UDP-N-acetyl-alpha-D-galactosamine:polypeptide N-acetylgalactosaminyltransferase 6 (GalNAc-T6)
220658_s_at	ARNTL2	56938		NM_020183	aryl hydrocarbon receptor nuclear translocator-like 2
220951_s_at	ACF	29974		NM_014576, NM_014576, NM_138932	apobec-1 complementation factor
221516_s_at	FLJ20232	54471		NM_019008	hypothetical protein FLJ20232
221653_x_at	APOL2	23780	607252	NM_030882, NM_030882	apolipoprotein L, 2
221920_s_at	MSCP	51312		NM_016612, NM_016612	mitochondrial solute carrier protein
222244_s_at	FLJ20618	55000		NM_017903	hypothetical protein FLJ20618

The minimum of three errors was found even using only 7 genes (Table 23).

Table 23. Genes used for the classification of MSS vs MSI tumors

Name	Symbol	Unigene	MSS	MSI
hepatocellular carcinoma-associated antigen 112	HCA112	Hs.12126	1261	653
metastasis-associated 1-like 1	MTA1L1	Hs.173043	45	91
chemokine (C-X-C motif) ligand 10	CXCL10	Hs.2248	104	274
heterogeneous nuclear ribonucleoprotein L	HNRPL	Hs.2730	194	630
hypothetical protein FLJ20618	FLJ20618	Hs.52184	776	388
splicing factor, arginine/serine-rich 6	SFRS6	Hs.6891	74	446
protein kinase C binding protein 1	PRKCBP1	Hs.75871	294	168

Classification of ambiguous samples

Application of the 7-gene classifier to the four samples showing ambiguity in the microsatellite analyses assigns all four to be microsatellite stable tumor class. Notably, all four showed expression levels of *Tumor Growth Factor β induced protein* (TFGBI), MLH1 and thymidylate synthase (TYMS) that are atypical for MSI tumors. Furthermore, these tumors were all from the left colon. Thus the misclassified tumors are clearly truly MSS or they belong to a yet undefined class of MSI tumors.

Stability of classification

To estimate the stability of the classifier based on all 97 tumor samples, we generated one hundred new classifiers based on randomly chosen datasets consisting of 30 MSS and 25 MSI samples. In each case the classifiers were tested with the remaining samples. The performance for each set was evaluated and averaged over all 100 training and test sets (Table 24). The mean error rate for MSS tumors was 0.52% and 1.38% for MSI tumors. The seven genes defined above were found to be those genes that were most frequently used in the crossvalidation loop. More than 50% of the errors were related to three tumors of which two were wrongly classified in all permutation and one in 94%. The remaining errors were mainly caused by four tumors with error rates of 40-47% showing that the former three samples are truly assigned contradictory to result from the microsatellite analysis and that four samples could not be assigned with confidence too any of the classes.

Table 24 Performance of the classifier

Trainings set	Test set
Errors in crossvalidation	Test errors

MSI	2.8% (n=25, range 0-6)	1.4% (n=10, range 0-4)
MSS	0.70% (n=30, range 0-3)	0.52% (n=29, range 0-2)
All	1.7% (n=55, range 1-7)	1.9% (n=39, range 0-5)

Table 25

Sensitivity, Specificity, and Predictive Value of Test for MSS based on the eight gene Classifier			
Positive for MSS	True = (0.9948*29)=28,8492		False = (0.138*10)= 1.38
Negative for MSS	False = (0.0052*29)= 0.1508		True = (0.962*10)= 9.62
Sensitivity	28.9507/29	=	99.5%
Specificity	9.62/10	=	96.2%
Positive predictive value	28.8492/30.2292	=	95.4%
Negative predictive value	9.62/9.7708	=	98.5%

*Based on a prevalence for MSS of 85%

Survival classifier

Using the same classification methods described above, we build classifiers for survival based on either all samples or the above defined groups of MSI-H and MSS. As seen in figure 10 a distinction of patient with good prognosis (>5 year survival) from patient with bad prognosis (< 5 years survival) can be achieved with higher precision and using only a fraction of the genes by first separating into MSI-H and MSS groups.

Construction of a classifier for sporadic versus hereditary microsatellite instable tumors

In order to identify a gene set for identification of hereditary microsatellite instable tumors we applied 19 sporadic microsatellite instable samples and 18 microsatellite instable samples to supervised classification as described above. We found ten genes we high scored for separation of sporadic MSI-H from hereditary MSI-H tumours (Table 26). In crossvalidation we found a minimum number of one error using two genes (Fig 9A) and were used in at least 36 of the 37 crossvalidation loops. The genes were: the mismatch repair gene MLH1 that show a general downregulation in sporadic disease and PIWIL1 that is lower expressed in hereditary cases (Fig 9B). Using these two genes only one error occurred: a sporadic microsatellite instable was classified as hereditary. Based on T-test we performed 500 permutations to test the significance of these two genes for marker genes and found both genes highly significant with p-values < 0.005.

Table 26

AFFYID	SYMBOL	LOCUSLINK	OMIM	REFSEQ	AFFYDESCRIPTION
206194_at	HOXC6	3223	142972	NM_004503	Homeo box C4
214868_at	PIWIL1	9271	605571	NM_004764.2	Piwi (Drosophila)-like 1
202520_s_at	MLH1	4292	120436	NM_000249.2	MutL (E. coli) homolog 1 (colon cancer, nonpolyposis type 2)
202517_at	CRMP1	1400	602462	NM_001313.2	Collapsin response mediator protein 1
205453_at	HOXB2	3212	142967	NM_002145.2	Homeo box B2 (HOXB2)
217791_s_at	PYCS/ADH18A1	5832	138250	NM_002860.2	Pyrroline-5-carboxylate synthetase (glutamate gamma-semialdehyde synthetase) (/PYCS/ADH18A1)
202393_s_at	TIEG	7071	601878	NM_005655.1	TGFB inducible early growth response (TIEG)
218803_at	CHFR	55743	605209	NM_018223.1	Checkpoint with forkhead and ring finger domains (CHFR)
219877_at	FLJ13842	79698		NM_024645.1	Hypothetical protein FLJ13842 (FLJ13842)
202241_at	C8FW	10221		NM_025195.2	Phosphoprotein regulated by mitogenic pathways (C8FW)

Cross platform classification

Real time PCR was applied both to verify the array data and examine if the 7-gene classifier would also perform on this platform. We chose 23 samples of which 18 were also analyzed on arrays. The correlation between the two platforms was high (data not shown). In order to test the performance of classification using PCR data we re-build our classifier with a 79 samples array dataset including only those tumors that were not analyzed with PCR. Two samples were classified in discordance with the microsatellite instability test of which one of them was ambiguously classified by the 7-gene array classifier.

Relation between microsatellite-instability status, stage and survival

Based on the 7-gene classifier, classification of 36 patients with Dukes' B tumors receiving no adjuvant chemotherapy, 18 were classified as MSI tumors and 18 as MSS tumors. The overall survival was highly significantly related to the classification since all nine patients that died within five years of follow-up were belonged to the

MSS group (P=0.0014) (Fig. 10A). Thus, the 7-gene classifier clearly proved to be a strong predictor of survival in Dukes B and it can be used to select patients who need adjuvant chemotherapy, namely those classified as MSS.

Among 65 patients with Dukes' C tumors receiving adjuvant chemotherapy, 17 were classified as MSI tumors and as 48 MSS tumors. Of these, 6 MSI and 27 MSS patients died within five years of follow-up meaning no significant difference in overall survival between these groups (P=0.55) (Fig. 10B). A trend was that the MSI showed a poorer short-term survival than the MSS, contrary to Dukes B patients. This difference can be attributed to the fact that a recent large study has shown that chemotherapy only benefit the MSS tumor patients, thus improving their survival to a level comparable to that which is characteristic of MSI tumor patients.

Clinical application of the discovery

In the clinic the 106 or less genes described can be used for predicting outcome of colorectal cancer when examined at the RNA level and also on the protein level as each gene identified is the project is transcribed to RNA that is further translated into protein. The genes can also be used determine which patient should be treated with chemotherapy as only non-microsatellite instable tumors will respond to 5-FU based therapy. Building classifiers can achieve a further stratification of patient with god and bad prognosis after stratification into microsatellite instable and stable tumors. The genes used to identify hereditary disease can be used to decide which patient should enter into sequencing analysis of mismatch repair genes.

The RNA determination can be made in any form using any method that will quantify RNA. The proteins can be measured with any method quantification method that can determine the level of proteins.

References

Agrawal D, Chen T, Irby R, Quackenbush J, Chambers AF, Szabo M, Cantor A, Coppola D, Yeatman TJ. Osteopontin identified as lead marker of colon cancer progression, using pooled sample expression profiling. J Natl Cancer Inst. 2002 Apr 3;94(7):513-21.

Birkenkamp-Demtroder K, Christensen LL, Olesen SH, Frederiksen CM, Laiho P, Aaltonen LA, Laurberg S, Sorensen FB, Hagemann R, ORntoft TF. Gene expression in colorectal cancer. Cancer Res. 2002 Aug 1;62(15):4352-63.

Boland CR, Thibodeau SN, Hamilton SR, Sidransky D, Eshleman JR, Burt RW, Meltzer SJ, Rodriguez-Bigas MA, Fodde R, Ranzani GN, Srivastava S. A National Cancer Institute Workshop on Microsatellite Instability for cancer detection and familial predisposition: development of international criteria for the determination of microsatellite instability in colorectal cancer. Cancer Res. 1998 Nov 15;58(22):5248-57. Review.

Chapusot C, Martin L, Bouvier AM, Bonithon-Kopp C, Ecarnot-Laubriet A, Rageot D, Ponnelle T, Laurent Puig P, Faivre J, Piard F. Microsatellite instability and intratumoural heterogeneity in 100 right-sided sporadic colon carcinomas. Br J Cancer. 2002 Aug 12;87(4):400-4.

Dyrskjot L, Thykjaer T, Kruhoffer M, Jensen JL, Marcussen N, Hamilton-Dutoit S, Wolf H, Orntoft TF. Identifying distinct classes of bladder carcinoma using microarrays. Nat Genet. 2003 Jan;33(1):90-6.

Frederiksen CM, Knudsen S, Laurberg S, Orntoft TF. Classification of Dukes' B and C colorectal cancers using expression arrays. J Cancer Res Clin Oncol. 2003 May;129(5):263-71.

Huang J, Qi R, Quackenbush J, Dauway E, Lazaridis E, Yeatman T. Effects of ischemia on gene expression.
J Surg Res. 2001 Aug;99(2):222-7.

Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res. 2003 Feb 15;31(4):e15.

Loukola A, Eklin K, Laiho P, Salovaara R, Kristo P, Jarvinen H, Mecklin JP, Launonen V, Aaltonen LA. Microsatellite marker analysis in screening for hereditary nonpolyposis colorectal cancer (HNPCC). Cancer Res. 2001 Jun 1;61(11):4545-9.

Markowitz S, Hines JD, Lutterbaugh J, Myeroff L, Mackay W, Gordon N, Rustum Y, Luna E, Kleinerman J.
Mutant K-ras oncogenes in colon cancers Do not predict Patient's chemotherapy response or survival. Clin Cancer Res. 1995 Apr;1(4):441-5.

Mori Y, Selaru FM, Sato F, Yin J, Simms LA, Xu Y, Olaru A, Deacu E, Wang S, Taylor JM, Young J, Leggett B, Jass JR, Abraham JM, Shibata D, Meltzer SJ. The impact of microsatellite instability on the molecular phenotype of colorectal tumors. Cancer Res. 2003 Aug 1;63(15):4577-82.

Ribic CM, Sargent DJ, Moore MJ, Thibodeau SN, French AJ, Goldberg RM, Hamilton SR, Laurent-Puig P, Gryfe R, Shepherd LE, Tu D, Redston M, Gallinger S. Tumor microsatellite-instability status as a predictor of benefit from fluorouracil-based adjuvant chemotherapy for colon cancer.
N Engl J Med. 2003 Jul 17;349(3):247-57.